



**Universidad Autónoma del Estado de México
Centro Universitario UAEM Valle de México**



Ingeniería en Computación

Unidad de Aprendizaje: Programación Paralela y Distribuida

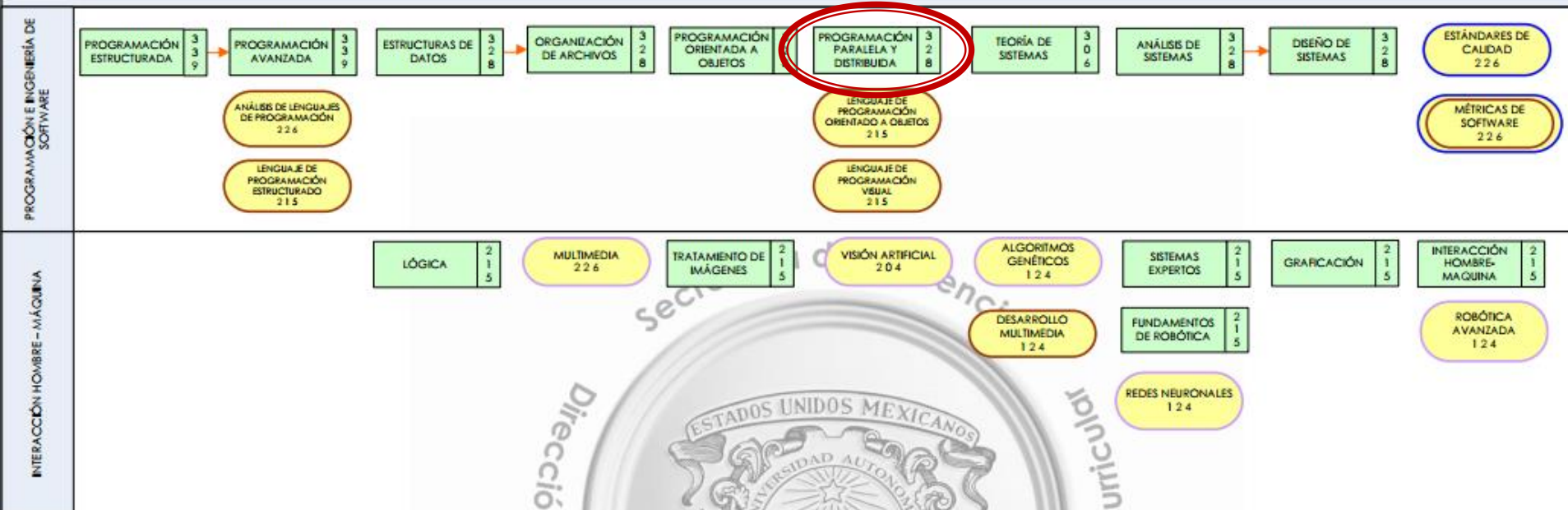
Tema: Introducción a los Sistemas Distribuidos

**Elaboró: Dr. en C. Héctor Rafael Orozco Aguirre
Octubre de 2016**

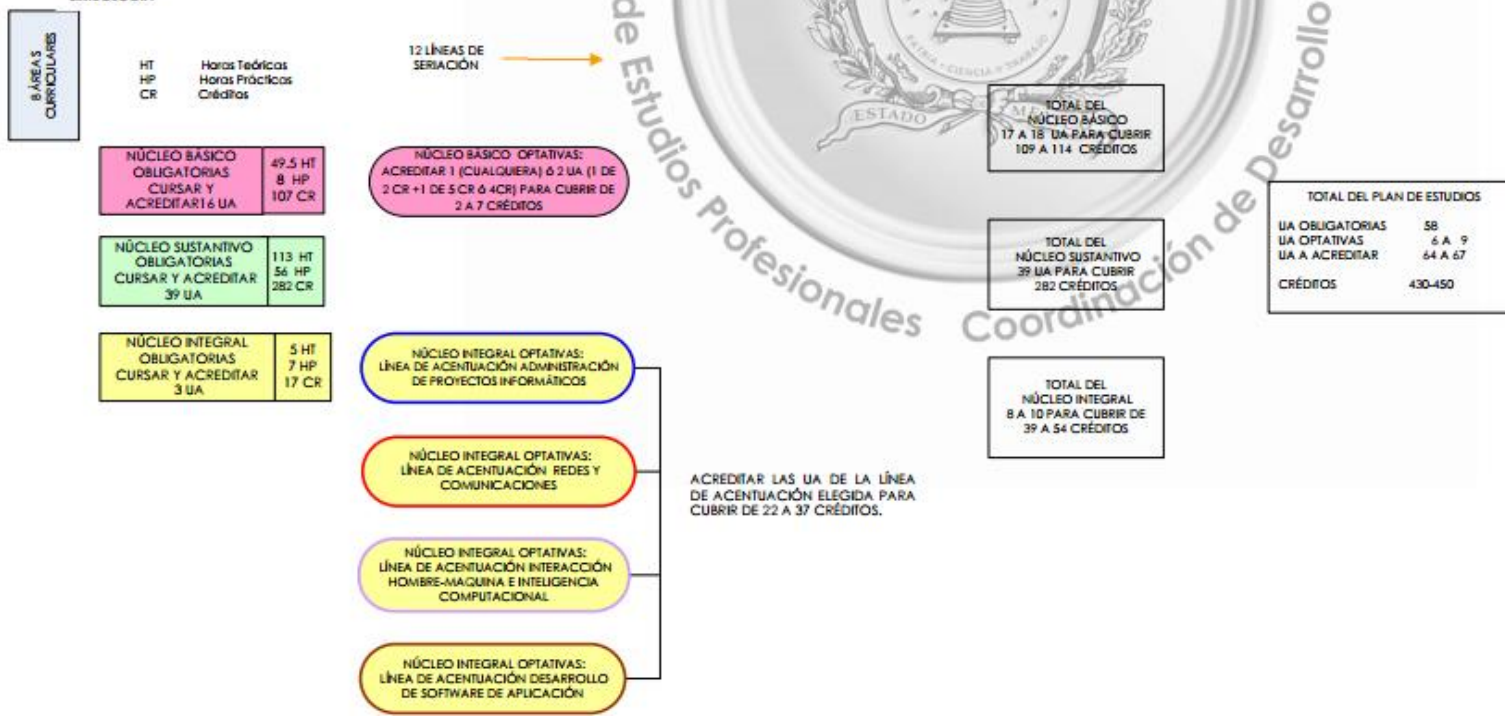
PROGRAMA DE ESTUDIOS POR COMPETENCIAS
PROGRAMACIÓN PARALELA Y DISTRIBUIDA

I. IDENTIFICACIÓN DEL CURSO

ORGANISMO ACADÉMICO: Facultad de Ingeniería						
PROGRAMA EDUCATIVO: Ingeniería en Computación				ÁREA DE DOCENCIA: Programación e Ingeniería del Software		
APROBACIÓN POR LOS H.H. CONSEJOS ACADÉMICO Y DE GOBIERNO		FECHA:		PROGRAMA ELABORADO POR: Ing. Luis E. Ledezma Fuentes Ing. Élfego Gutiérrez Ocampo		PROGRAMA REVISADO POR: Integrantes de la Academia de Programación e Ingeniería de Software
				FECHA DE ELABORACIÓN : Enero 2011		FECHA DE REVISIÓN : Mayo 2011
CLAVE	HORAS DE TEORÍA	HORAS DE PRÁCTICA	TOTAL DE HORAS	CRÉDITOS	TIPO DE UNIDAD DE APRENDIZAJE	NÚCLEO DE FORMACIÓN
L41057	2	3	5	8	Curso-Laboratorio	Sustantivo
PRERREQUISITOS: Programación Avanzada, Estructuras de Datos		UNIDAD DE APRENDIZAJE ANTECEDENTE: Ninguna		UNIDAD DE APRENDIZAJE CONSECUENTE: Ninguna		
PROGRAMAS EDUCATIVOS O ESPACIOS ACADÉMICOS EN LOS QUE SE IMPARTE: Licenciatura en Ingeniería en Computación (Facultad. de Ingeniería, Centros Universitarios: Atlacomulco, Ecatepec, Texcoco, Valle de Chalco, Valle de México, Valle de Teotihuacán, Zumpango)						




SIMBOLOGÍA



Propósito de la Unidad de Aprendizaje

- ▶ Presentar al alumno con la tecnología de vanguardia en el diseño de algoritmos y programación en paralelo haciendo uso de computadoras con 2 o más procesadores o haciendo uso de procesadores de doble núcleo con vistas a capacitar al estudiante a su egreso en el análisis, diseño, desarrollo y construcción de sistemas de resolución de problemas de gran envergadura en donde se requiera el uso de arquitecturas paralelas o sistemas distribuidos.
 - Analizar y diseñar algoritmos y programas haciendo uso de técnicas de programación en paralelo aplicables a la tecnología computacional.
 - Analizar y diseñar proyectos donde se requiera programación en paralelo y distribuida.
 - Comunicarse con expertos de otras áreas.
 - Utilizar eficazmente computadoras con 2 o más procesadores.
 - Analizar soluciones del entorno y problemas propios de ser tratados mediante sistemas computacionales paralelos y distribuidos.
 - Proponer soluciones eficaces y eficientes.
 - Crear nuevas ideas para la solución de problemas.
 - Aplicar los conocimientos mediante prácticas.
 - Conocer la temática básica de la profesión que desempeña en la práctica.
 - Especificar arquitecturas de computadoras de arquitectura paralelas o sistemas distribuidos

Contenido

- Hechos relevantes y consecuencias de uso
 - Definición
 - Concurrencia y paralelismo
 - Ventajas y desventajas de uso
 - Características
 - Problemas a considerar
- 

Guion explicativo

- ▶ Esta presentación tiene como fin dar a conocer a los alumnos los siguientes aspectos:
 - ¿Qué hechos dieron origen a los sistemas distribuidos?
 - Definición, ventajas y desventajas de uso de un sistema distribuido
 - Características más destacadas de los sistemas distribuidos
 - Principales problemas a ser tomados en cuenta

Guion explicativo

- ▶ El contenido de esta presentación contiene temas de interés contenidos en la Unidad de Aprendizaje de Programación Paralela y Distribuida.
- ▶ Las diapositivas deben explicarse en orden, y deben revisarse aproximadamente en 2 horas, además de realizar preguntas a la clase sobre el contenido mostrado.

Hechos relevantes

- ▶ En los 80s ocurren dos hechos importantes:
 - Se hace disponible la tecnología de microprocesadores (aparecen las microcomputadoras).
 - Los sistemas distribuidos desarrollan los sistemas de redes.
- ▶ Finalmente, durante los 90s ambos sistemas de comunicaciones y sistemas de computación hacen frente a las necesidades de procesamiento y distribución tanto de información como de procesamiento.

Consecuencias de uso

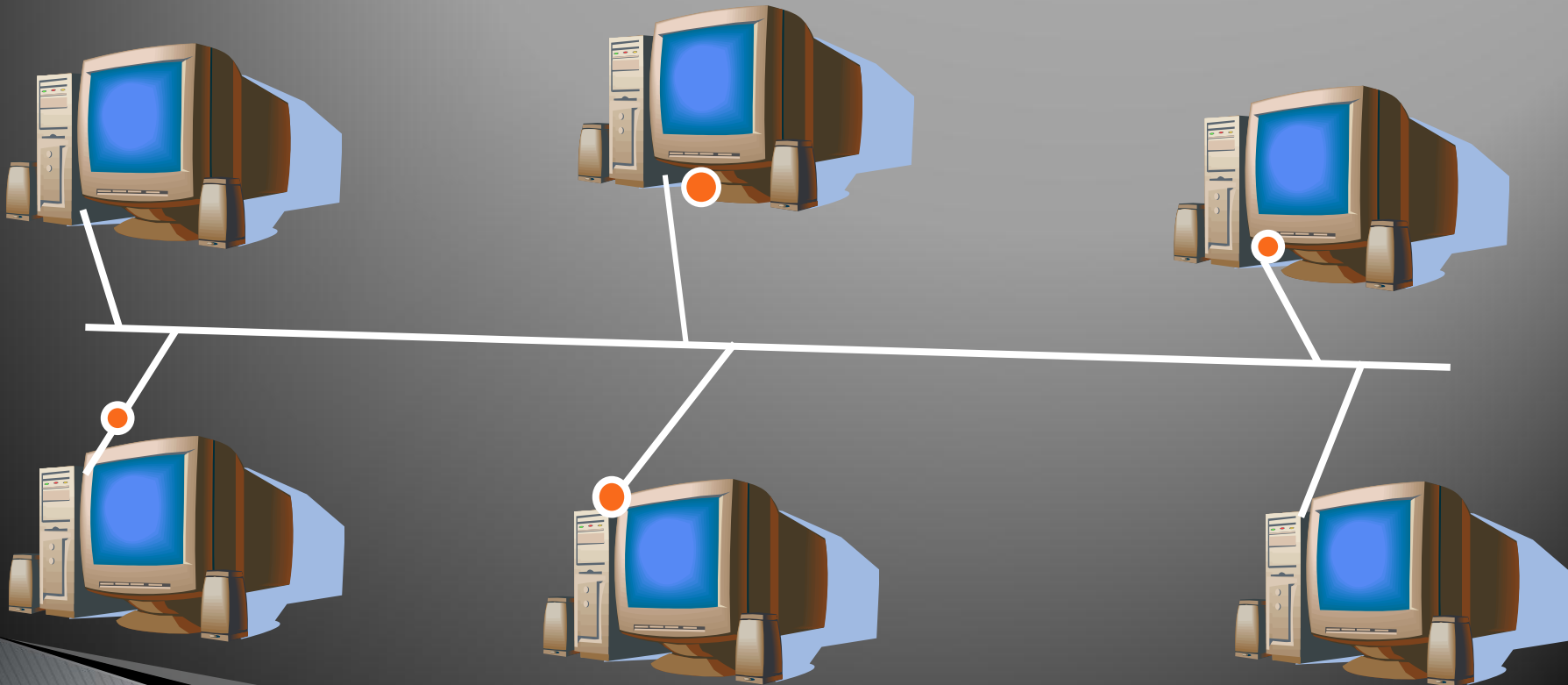
- ▶ Fácil y económico poner a trabajar un conjunto de maquinas a trabajar para resolver un problema.
- ▶ Acelerar la velocidad de respuesta del sistema.
- ▶ Necesidad de sincronizar el trabajo de cada componente del sistema ☹.

Definición técnica

Componentes de Hardware y Software

Unidos mediante una red

Comunican y coordinan sus acciones mediante el paso de mensajes



Ventajas de uso

- ▶ Relativamente fácil y económico implementar un sistema compuesto de un conjunto de procesadores conectados por medio de una red.
- ▶ Varios procesadores trabajan más rápido que uno solo.
- ▶ Mejor adaptado a la organización de la sociedad humana.
- ▶ Fiabilidad o robustez.

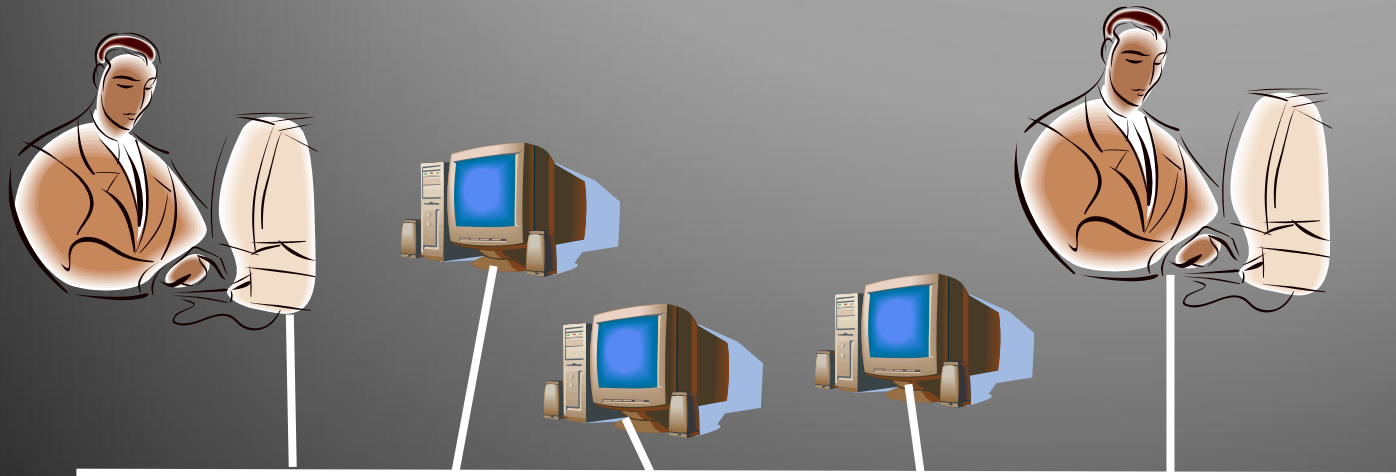
Ventajas de uso

- ▶ Disposición de información en todo el sistema.
- ▶ Extensibilidad
- ▶ Posibilidad de compartir recursos escasos.
- ▶ Buena relación costo y eficiencia.

Concurrencia y paralelismo

Yo puedo realizar mi trabajo en mi computadora mientras tú lo realizas en tu computadora.

La capacidad del sistema para manejar más recursos compartidos se puede incrementar añadiendo más computadoras.

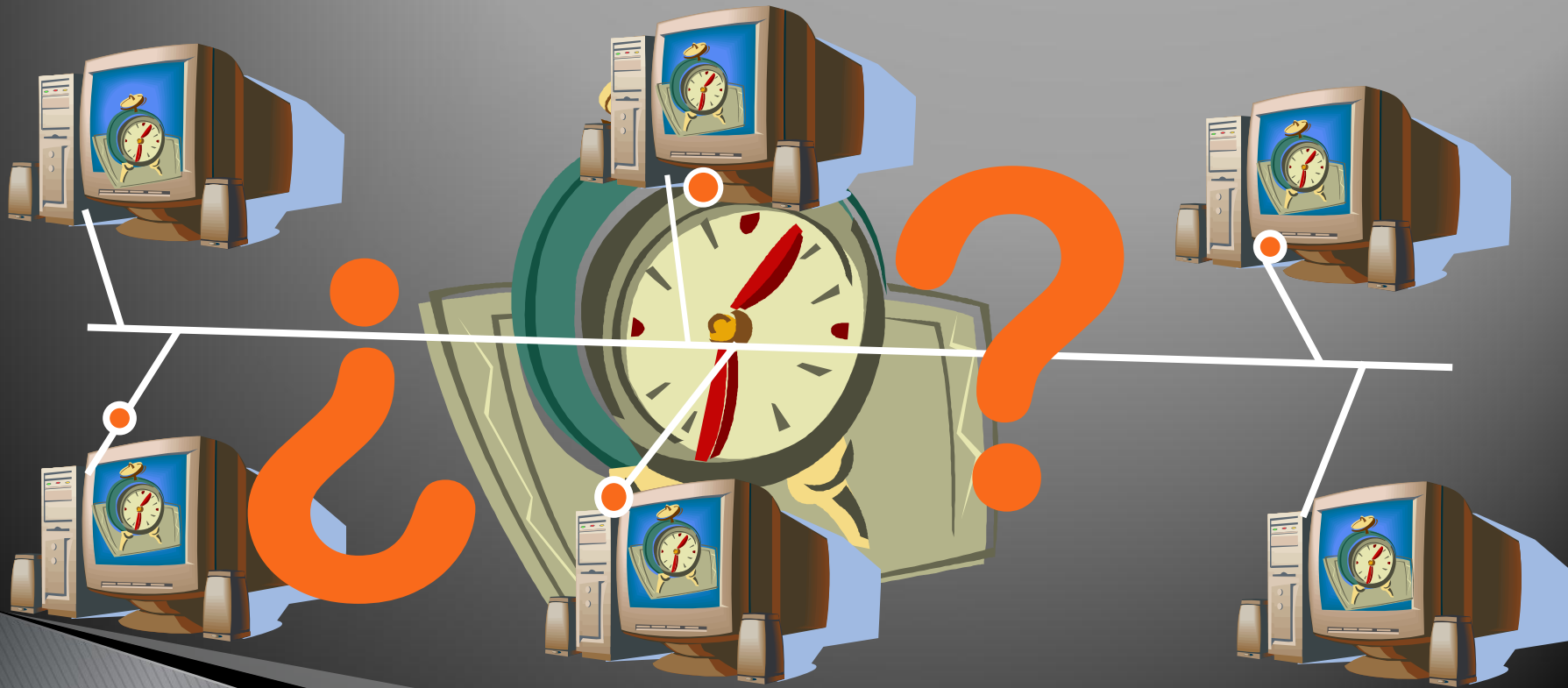


Desventajas de uso

- ▶ Imposibilidad de tener un sistema de comunicaciones ideal con cero tiempo de retardo. Esto es lo que dificulta todos los problemas comunes de los sistemas distribuidos:
 - Robustez, interbloqueo, seguridad, etc.
 - Imposible de establecer un estado global del sistema.

Inexistencia de un reloj global

Los programas coordinan sus acciones mediante un sistema de paso de mensajes.
Hay límites de precisión en los relojes de los computadores que impide sincronizar su reloj.
No hay una posición global del tiempo correcto.



Problema ejemplo

Juan y Pedro están en dos bares, uno enfrente del otro.

Afuera hay un ratero que espera a que salga uno de los dos solos, pero si salen al mismo tiempo, no los asaltará.

Ellos saben de eso, pero la única forma en la que pueden comunicarse y ponerse de acuerdo es a través de mensajes en su celular.



¿Qué sucede con los mensajes del celular?

- ▶ Pueden llegar casi inmediatamente.
- ▶ Pueden retardarse:
 - El usuario destino no tiene recepción.
 - Tiene su teléfono apagado.
 - Falla en el sistema de comunicaciones.
- ▶ Pueden perderse.



Características

- ▶ Uso de un sistema de comunicación.
- ▶ Ausencia de memoria común.
- ▶ Sincronización del trabajo.
- ▶ Imposibilidad de que un observador pueda determinar un estado global.
- ▶ Comunicación a través de mensajes.

Problemas a considerar

- ▶ Sistemas de comunicación no fiables:
 - es decir posibilidad de perdida duplicación o corrupción de mensajes.
- ▶ Comunicación insegura:
 - lo que se traduce en la posibilidad de que un agente externo intercepte o afecte la información entre las entidades del sistema.

Problemas a considerar

- ▶ Costo de comunicaciones:
 - implica reducción del ancho de banda, mayor latencia y altos costos de comunicación entre procesos independientes.
- ▶ Fallas locales:
 - no deben impedir realizar el trabajo deseado.

Redes

- ▶ Problemas (WAN)
 - Confiabilidad
 - Selección de rutas de comunicación
 - Control de tráfico
 - Prevención de cuellos de botella
 - Seguridad

Redes

- ▶ Problemas LAN
 - Difusión y Sincronización
 - Elección
 - Detección de terminación
 - Asignación de recursos
 - Exclusión Mutua
 - Interbloqueo
 - Mantenimiento de archivos distribuidos

Algoritmos

- ▶ Algoritmos Distribuidos
 - Implementación de un protocolo de envío y recepción de mensajes
 - Implementación de memoria virtual compartida
 - Balance de carga
 - Tolerancia a fallas

Distribución

- ▶ ¿ Qué es lo que se distribuye?
 - Información (datos).
 - Control (procesamiento).

Distribución de datos

- ▶ Posibilidades para efectuar la distribución:
 - Duplicación de información.
 - Fragmentación de la información.

Duplicación de la información

Duplicar consiste en distribuir copias completas de la información

Los problemas de la duplicación son:
como asegurar la *consistencia*
de la información. Asegurar una
Distribución optima,etc..

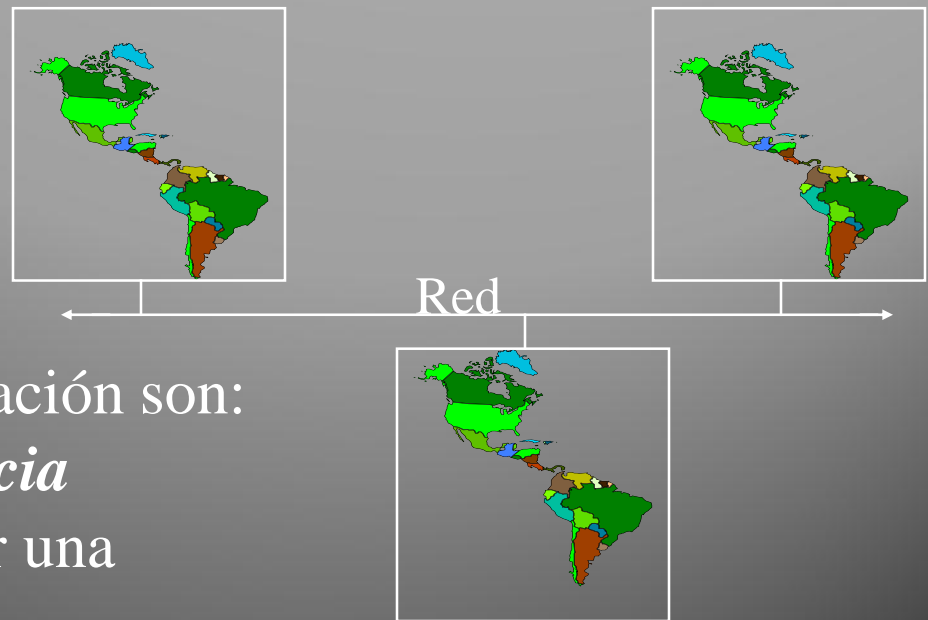


Fig.. Duplicación de información

Fragmentación de la información

La *fragmentación* consiste en dividir la información en segmentos más “fáciles” de manejar.

Existen varios problemas que deben de ser resueltos. Por ejemplo: Fragmentación apropiada, localización de los fragmentos, etc..

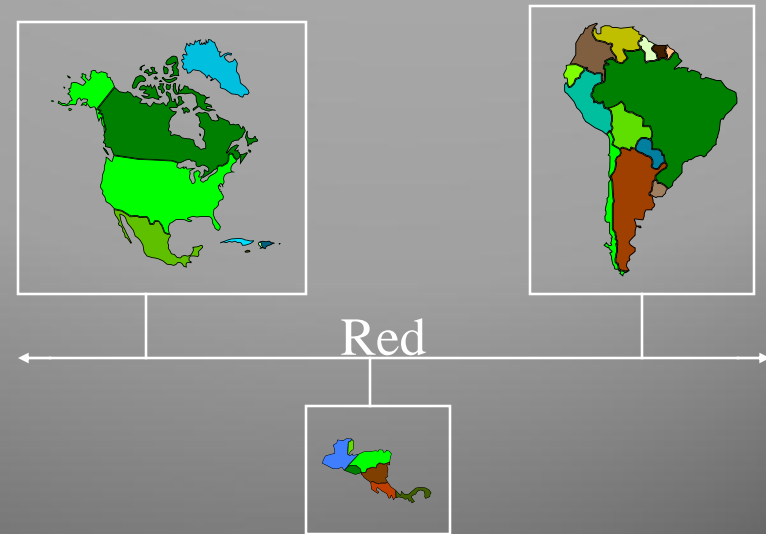


Fig. Información fragmentada

Distribución del control

- ▶ En un sistema distribuido, no debe existir una sola entidad que dirija el sistema. Es decir, no existe una jerarquía estática
- ▶ Puede ser necesario un algoritmo de asignación de funciones.
- ▶ En general, el control depende de la aplicación (tipo de sistema).

Áreas de aplicación

- ▶ Usos más relevantes:
 - Telecomunicaciones.
 - Diseño de protocolos.
 - Procesamiento distribuido.
 - Cálculo científico.
 - Procesamiento tiempo real.
 - Sistemas operativos.
 - Bases de datos.

Problemas más comunes

- ▶ Tolerancia a fallas
- ▶ Confiabilidad
- ▶ Consistencia

Tolerancia a fallas

- ▶ El sistema distribuido es vulnerable a la falla de algunos de sus elementos.
- ▶ El sistema debe seguir funcionando aunque sea con un menor desempeño.
- ▶ Debe contar con módulos para:
 - Detección de errores.
 - Evaluación y recuperación de daños.
 - Reparación y continuidad del servicio.

Confiabilidad

- ▶ Se refiere a la manera en que los datos son transmitidos a través de las vías de comunicación.
- ▶ La confiabilidad debe tratar con la posibilidad de pérdida y modificación de los datos.

Consistencia

- ▶ Diferentes procesos pueden acceder y actualizar los mismos datos concurrentemente.
- ▶ Aunque los cambios no son instantáneos cualquier par de procesos debe percibir el mismo estado de los datos.

Algunos conceptos

► Proceso

- Término introducido por Dijkstra en 68 para modelar las relaciones entre las diferentes unidades de ejecución que deben compartir recursos comunes.
- En los sistemas distribuidos un proceso representa la unidad de ejecución elemental de un algoritmo distribuido o paralelo.
- Varios procesos pueden ejecutarse simultáneamente. Se habla de procesos secuenciales cuando existe un solo flujo de control

Algunos conceptos

- ▶ Las vías de comunicación son el medio a través del cual viajan los mensajes.
- ▶ Las vías de comunicación tienen básicamente dos propiedades:
 - Propiedades estructurales.
 - Propiedades comportamentales.

Vías de comunicación

► Propiedades estructurales:

- Se refiere a las mallas de comunicación.
- Todas las topologías son posibles
 - Cada una tiene sus características que se reflejan en los algoritmos distribuidos que las resuelven.
- Las topologías más comunes
 - El anillo, la estrella, el árbol y el grafo completo.

Vías de comunicación

- ▶ Propiedades comportamentales:
 - Se refieren al comportamiento de las vías de comunicación que queda especificado por:
 - Transmisión con o sin duplicación.
 - Transmisión sin alteración de mensajes.
 - Orden de emisión y recepción de mensajes.
 - Tiempo de espera de un mensaje finito (sin perdida).
 - Tiempo de espera acotado (con perdida o sin perdida).

Problemas a estudiar

- ▶ La repartición y distribución de recursos de cálculo y de almacenamiento.
- ▶ La exclusión mutua.
- ▶ El interbloqueo, etc.